ARTICLE

A Method for Mapping Glycosylation Sites in Proteins

Mark S. Han[1], and John T. Simpson[2]*

[1]*Reference Standards Laboratory and* [2]*Compendial Development Laboratory, The United States Pharmacopeial Convention, Rockville, Maryland 20852, USA*

The analysis of protein glycosylation by mass spectrometry (MS) has been a challenging technical problem. Quantification by HPLC of *N*-linked glycans can be executed by the use of peptide-*N*-glycosidase F to release them from the protein, followed by attachment of a fluorescent label and subsequent fluorescence detection. Similar quantification of *O*-linked glycans is not possible, as a result of the lack of a universal deglycosylation enzyme. Site-specific analyses by MS, such as the use of proteases to digest the glycoprotein, are difficult to use for quantification of glycans, as a result of the presence of miscleavages. Here, we present a new application of a digestion method for native proteins using resin-bound, thermally stabilized proteases. The use of this enzymatic treatment eliminates miscleavages around the site of glycosylation, thereby allowing site-specific relative quantification of glycans on glycoproteins. A native, intact human mAb was digested using a thermally stable, resin-bound trypsin to produce glycopeptides from the Fc region using a single-step protocol. A 1 mg sample was treated with 60 μg trypsin for 3 h at 70°C. After digestion, acetonitrile was added, and the mixture was centrifuged to remove the resin before analysis. Liquid chromatography (LC)/MS with hydrophilic interaction chromatography was used to analyze the glycopeptides produced. All of the glycopeptides found resulted from a single peptide (EEQYNSTYR). The LC/MS analysis of the glycopeptides is compared with that of fluorescently labeled glycans. Quantitative analysis produced a correlation coefficient of 0.87 for the linear fit between the glycopeptide and released glycan methods.

KEY WORDS: glycans, LC/MS, quantification, site-specific glycosylation

INTRODUCTION

Protein glycosylation is an important post-translational modification that has been shown to affect stability, conformation, and activity of glycoproteins.[1] In biotherapeutics, its effects on pharmacokinetics and efficacy have caused glycosylation to become a critical quality attribute that is engineered for optimal performance.[2] Many complementary approaches for the analysis of protein glycosylation have been developed, but there is no single method that provides site-specific localization and quantification.[3–5]

The analysis of *N*-linked glycans is often carried out by releasing them from the protein using peptide *N*-glycosidase F (PNGase F), followed by fluorescent labeling and HPLC separation with fluorescent detection. Although the availability of a single enzyme that can cleave almost all glycans

from asparagine greatly simplifies the analysis of *N*-linked glycans, release of the glycans from the protein destroys the data for site specificity. In addition, different glycans are released at different rates by PNGase F, presumably as a result of steric effects. Furthermore, release of the same glycans from different proteins can vary in speed. These issues, among others, must be taken into account before PNGase F is used as a universal deglycosylation agent in quantification.[6]

O-Linked glycosylations are more problematic as a result of the lack of an equivalent enzyme to PNGase F for these linkages. Removal of *O*-glycans can be facilitated by chemical cleavage and some enzymes.[3] For proteins such as erythropoietin that have a single *O*-glycosylation site and multiple *N*-glycosylation sites, PNGase F has been used to remove the *N*-glycans, and *O*-glycans have been analyzed using intact protein mass measurement.[7] However, assignment of the location of glycosylation is difficult using this approach.

To determine site specificity (localization) of glycosylations, proteases are often used.[8] However, the use of the standard reduction, alkylation, and digestion methodology does not yield perfect efficiency; therefore, missed cleavages occur. A recent study on trypsin by Šlechtová *et al.*,[9] details some of these issues. Miscleavages cause complications by a multiplicative factor in that the total number of glycopeptides generated will equal N(M + 1); where N is the number

*ADDRESS CORRESPONDENCE TO: Mark S. Han, Reference Standards Laboratory, USP, 12601 Twinbrook Parkway, Rockville, MD 20852, USA. E-mail: msh@usp.org

of glycans, and M is the number of miscleavages around the desired peptide fragment. Miscleavages also dilute each individual glycan with their occurrences. The key to use of glycopeptides for quantification of glycosylations is to minimize miscleavages.

Recently, thermally stable, resin-bound proteases (SMART Digest; Thermo Fisher Scientific, Waltham, MA, USA) have been introduced. Along with the advantages of using a solid support for the enzymes, the thermal stability allows digestions to occur at elevated temperatures, which denature the substrate protein without affecting the protease.[10] Therefore, even proteins, such as albumin, which is difficult to digest without reduction and alkylation of its 17 disulfide bridges, can be digested in the native state (Meyer et al.,[11] poster; http://www.perfinity.com/flash-digest-applications-by-type). This greatly simplifies the workflow for digestion. The efficiency of reduction/alkylation and its effect on the protein need not be considered when using the SMART Digest procedure. In addition, the use of temperature ranges outside of normal in-solution trypsin digestions with SMART Digest offers various experimental opportunities.

To evaluate the application of the SMART Digest protocol to the determination of protein glycosylation, the National Institute of Standards and Technology (NIST; Gaithersburg, MD, USA) mAb reference material was used as a model protein.[12] Because it is a human mAb with a single point of glycosylation at asparagine 297 in the Fc region, glycopeptides generated by its digestion could be used to compare most mAb biotherapeutics. Furthermore, if digestion of the NIST mAb could be optimized to minimize miscleavages around the site of glycosylation, then the protocol could also be applied to other human mAb for monitoring this critical quality attribute.

In the best-case scenario, with acceptable digestion efficiency, the tryptic peptide at asparagine 297, EEQYNSTYR, could form the basis for an isotopically labeled glycopeptide standard. Relative quantification, *via* lot-to-lot comparisons of a particular biotherapeutic/biosimilar, could be obviated by use of the appropriate standards. Absolute quantities of particular glycans at a specific site of glycosylation could be measured along with their historical variations, thereby giving increased confidence with respect to acceptable glycosylation variance for the biotherapeutics.

MATERIALS AND METHODS

Glycoworks RapiFluor-MS N-Glycan Starter Kit-24 samples were purchased from Waters (Milford, MA, USA). Milli-Q Advantage A10 and Amicon Ultra 10K molecular weight cutoff filters were obtained from Merck Chemicals GmbH (Darmstadt, Germany). Ammonium formate was purchased from Fluka Chemical (Milwaukee, WI, USA). Buffer-free SMART Digest solution was provided by Perfinity Biosciences (West Lafayette, IN, USA; currently supplied by Thermo Fisher Scientific). Formic acid and HEPES were obtained from Sigma-Aldrich (St. Louis, MO, USA). Axygen MaxyClear Snaplock microtubes, 1.5 ml, were obtained from Fisher Scientific (Pittsburgh, PA, USA). Acetonitrile, pH 3.00 buffer, pH 6.00 buffer, SMART Digest trypsin kit, and sodium hydroxide solution were purchased from Thermo Fisher Scientific. The NIST mAb was purchased from NIST. Purified water was produced from in-house Milli-Q systems.

Solution preparation

HEPES SMART Digest buffer was prepared as follows: buffer-free SMART Digest solution, 70 ml, and HEPES, 476 mg, were added to a 250 ml bottle and mixed until fully dissolved. The solution was adjusted to pH 7.7 using 5 N sodium hydroxide.

Sample preparations

Samples were prepared as described.

NIST mAb in HEPES SMART Digest buffer

NIST mAb, 0.5 ml, was added to an Amicon 10K molecular weight cutoff filter and centrifuged at 14,000 g for 15 min. An additional 0.3 ml NIST mAb was added to the retentate, and the sample was centrifuged at 14,000 g for 15 min. HEPES SMART Digest buffer, 0.5 ml, was added, and the sample was centrifuged at 14,000 g for 15 min. The buffer exchange was repeated twice more. The sample was collected at 1000 g for 5 min and diluted to 1.6 ml using additional HEPES SMART Digest buffer. The final concentration of the sample was 1 mg mAb/200 μl.

TABLE 1

HPLC gradient program

Time, min	%B	Flow rate, ml/min
0	75	0.4
1	75	0.4
35	50	0.4
36.5	0	0.2
39.5	0	0.2
43.1	75	0.2
47.6	75	0.4
55	75	0.4

Mobile phase A: 100 mM ammonium formate (pH 4.4). Mobile phase B: acetonitrile. %B, Percent of mobile phase B at each gradient step.
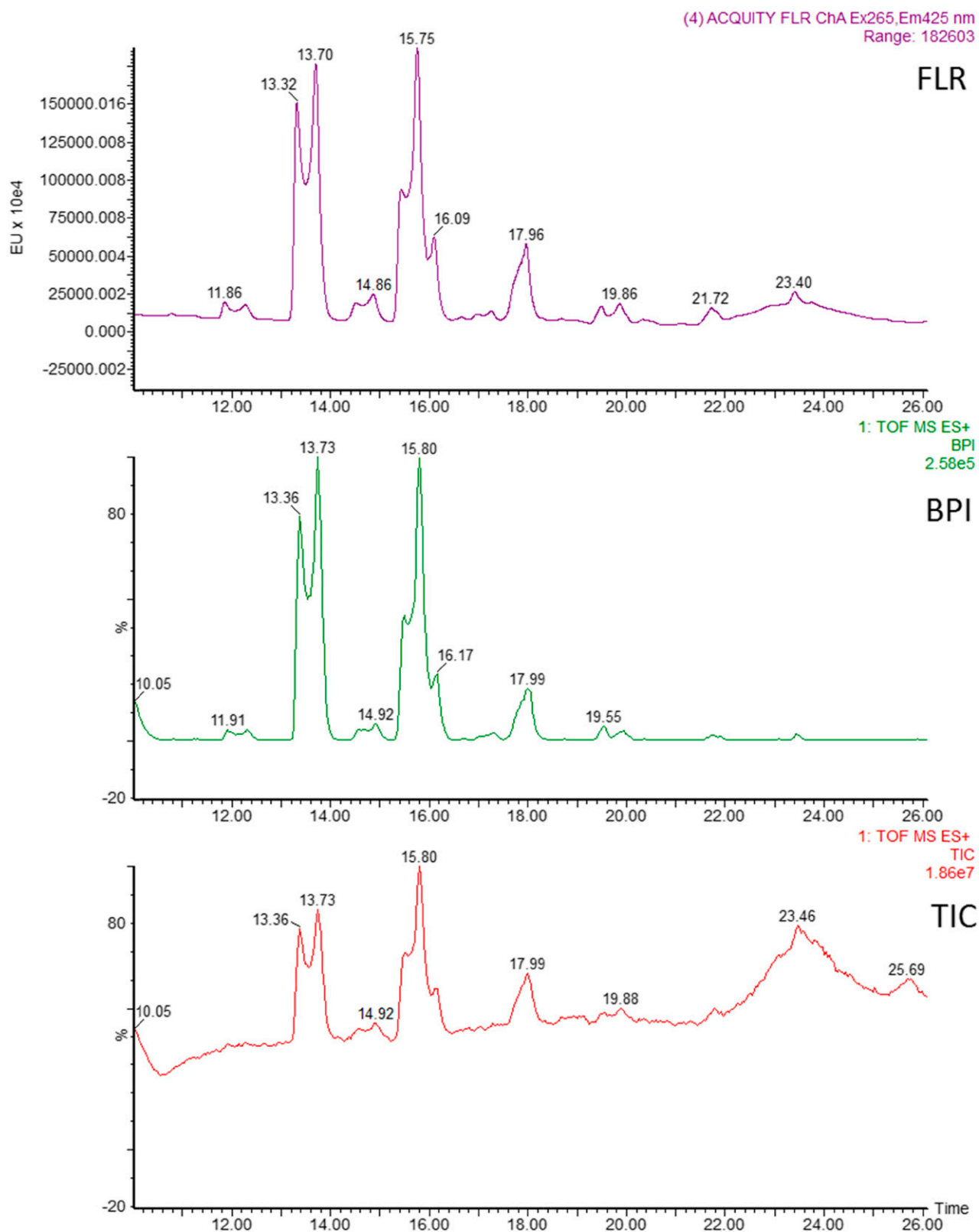
FIGURE 1

Elution profile of RapiFluor-labeled PNGase F-released glycans from NIST mAb. Note that the glycans are observed in the FLR, the base peak intensity (BPI) ion, and the TIC chromatograms. Expansion of the chromatograms is shown for the FLR, BPI, and TIC for clarity. EU, emission units; TOF, time of flight; ES+, positive electrospray ionization.
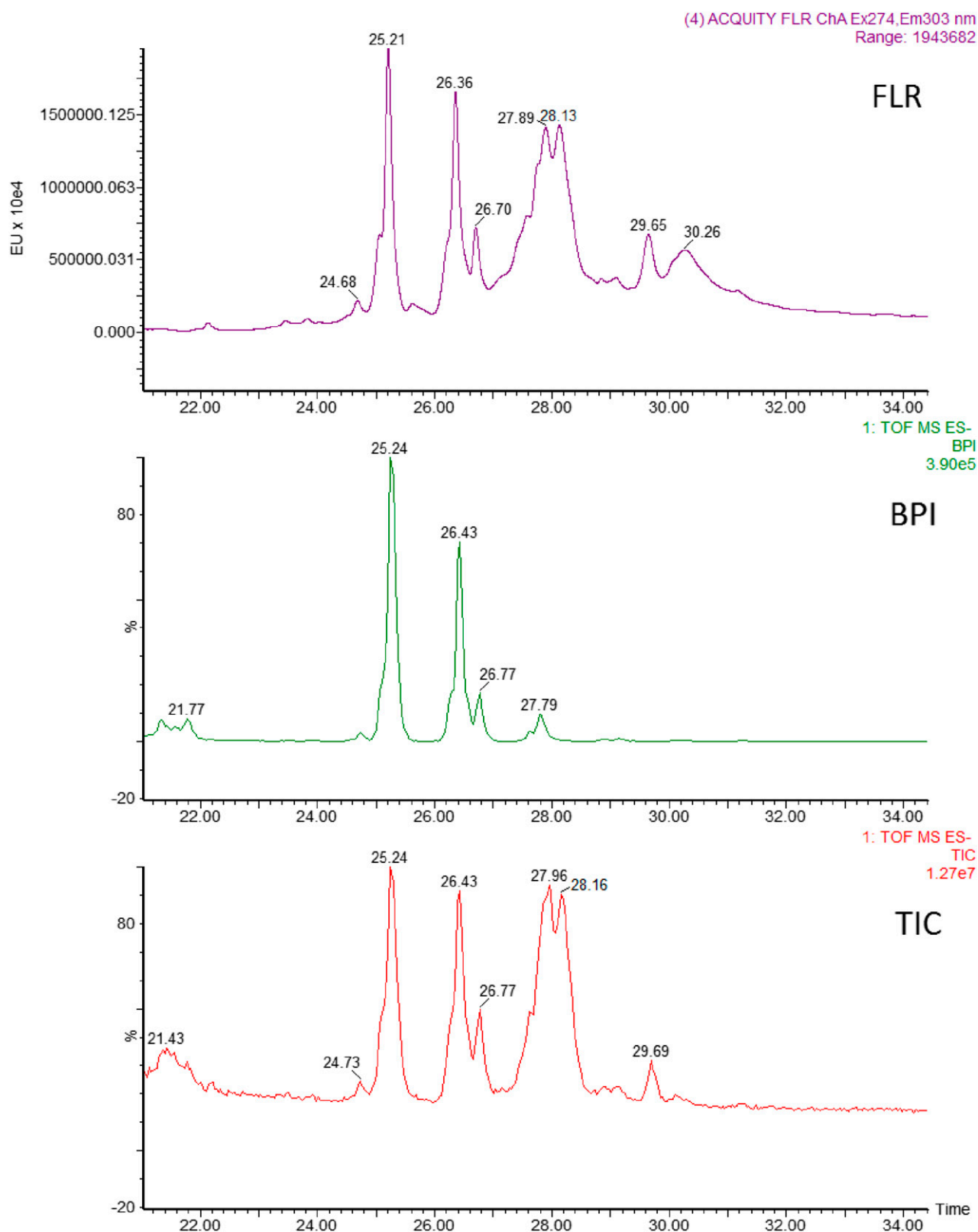
FIGURE 2

Elution profile of glycopeptides obtained from the digestion of native NIST mAb. Note the FLR monitors the presence of tyrosine in the tryptic peptide EEQYNSTYR. The glycopeptides are observed in the FLR trace, the base peak intensity (BPI) ion, and the TIC chromatograms of the mass spectrometer. Expansion of the area of the chromatogram containing the eluting glycopeptides is shown for clarity. ES−, Negative electrospray ionization.

NIST mAb glycopeptides

NIST mAb in HEPES SMART Digest buffer, 200 μl, was added to a microtiter well containing resin-bound SMART Digest trypsin, ~20 μg enzyme. The resin was suspended in the sample, and the contents were transferred to another microtiter well containing a fresh aliquot of resin-bound trypsin. The resin was suspended in the sample, and the contents were transferred to a third microtiter well containing a fresh aliquot of resin-bound trypsin. This yielded a 3-fold increase in the trypsin available to digest the sample. The sample was digested on a digital shaking drybath at 70°C for 3 h at 1400 rpm. The digest mixture was diluted with 200 μl acetonitrile and centrifuged at 21,100 *g* for 15 min. The resulting supernatant was collected for analysis.

NIST mAb RapiFluor-labeled glycans

Water, 15.3 μl, was added to a 1.5 ml microcentrifuge tube. NIST mAb in HEPES SMART Digest buffer, 5 μl, was also added followed by 8.5 μl additional HEPES SMART Digest buffer. The mixture was incubated at 90°C for 3 min in a dry block bath. The sample was allowed to cool to room temperature. Rapid PNGase F, 1.2 μl, was added to the sample, followed by incubation at 50°C for 30 min at 600 rpm on a ThermoMixer R (Eppendorf, Hauppauge, NY, USA). *N,N*-Dimethylformamide, 100 μl, was added to 8 mg RapiFluor powder and mixed to produce the RapiFluor solution, 12 μl of which was added to the sample. The resulting solution was vortexed and allowed to sit at room temperature for 15 min. Acetonitrile, 360 μl, and water, 240 μl, were added to the sample. The sample was mixed by vortexing until clear.

HILIC separation of glycopeptides and labeled glycans

Chromatography was performed on a Waters 2D Acquity UPLC equipped with photodiode array and fluorescence (FLR) detectors and interfaced with a Waters Synapt G2-S mass spectrometer (see MS conditions). A hydrophobic interaction chromatography (HILIC) mode of separation was used with a Waters UPLC Glycan BEH Amide column, 150 × 2.1 mm, 1.7 μ. The column temperature was set to 60°C, and the autosampler temperature was set to 5°C. The injection volume was 50 μl. The photodiode array scan range was 190–700 nm. The FLR was set to excitation 265 nm, emission 425 nm for RapiFluor-labeled glycans and excitation 274 nm, emission 303 nm for tyrosine present in the glycopeptides.

The initial flow rate was 0.4 ml/min with mobile phase A consisting of 100 mM ammonium formate (pH 4.4) and
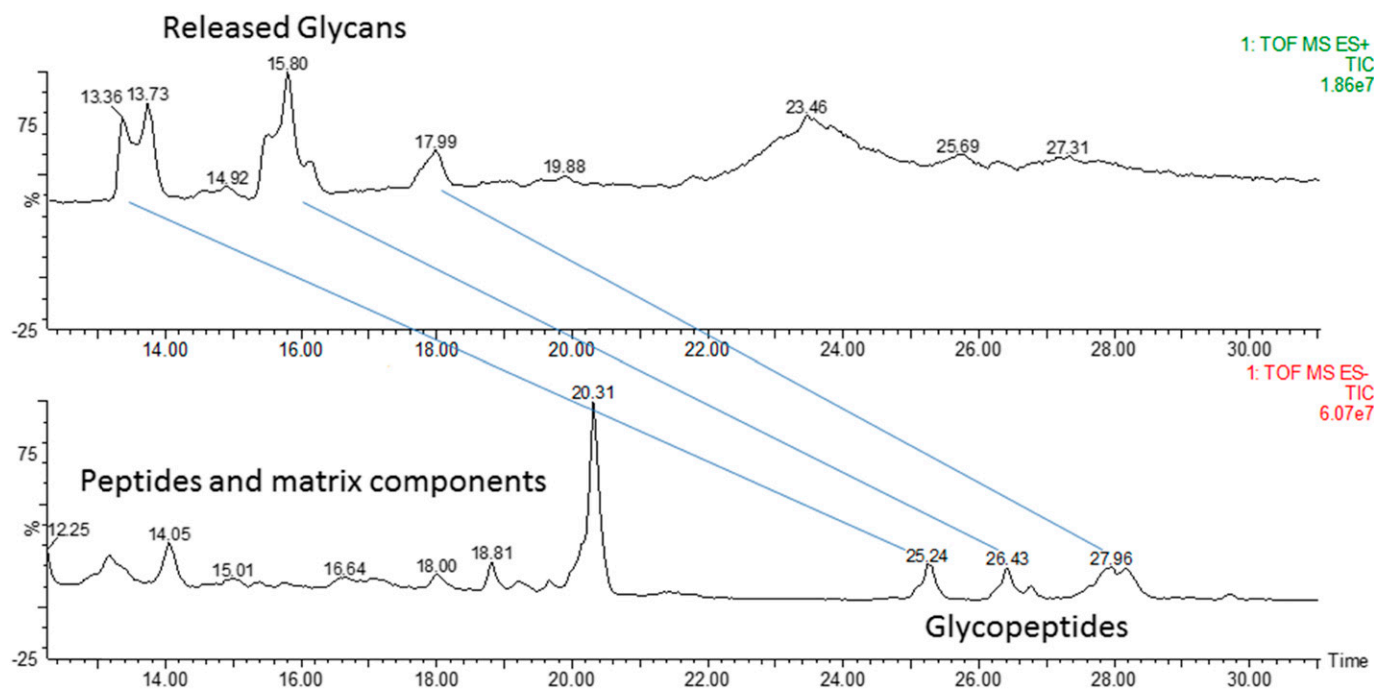
FIGURE 3

Comparison of TIC chromatograms for PNGase F-released NIST mAb glycans. *Upper*) Chromatogram and glycopeptides; *lower*) chromatogram from treatment of the NIST mAb with the SMART Digest protocol. Note that the glycopeptides are more strongly retained by the HILIC separation method. Peptides and other matrix components present in the native digestion of NIST mAb are seen eluting earlier in the glycopeptide TIC. As the other components in the glycopeptides sample are easily resolved from the desired analytes, extensive sample cleanup is not necessary.

mobile phase B being acetonitrile. The full gradient is detailed in **Table 1**.

MS conditions

Liquid chromatography/mass spectrometry (LC/MS) experiments were conducted on a Waters Synapt G2-S mass spectrometer. The scan range was mass-to-charge ratio 100–2400 for positive and negative ion mode analyses. Scan time was 1 s, and glu-fibrinopeptide B was constantly infused (2 μl/min) as a calibrant ("lock mass"). The capillary voltage was set to 2.5 kV, with a source temperature of 120°C and desolvation temperature of 500°C. The nitrogen nebulizer gas flow was set to 700 l/h.

RESULTS

HPLC analysis of the PNGase-released, fluorescently labeled glycans from the NIST mAb produced a complex chromatogram with overlapping components. A comparison of the FLR chromatogram and LC/MS total ion current (TIC) chromatogram of the region of interest is presented in **Fig. 1**. Glycans were identified by their mass and elution order in the HILIC chromatography method using the glucose unit value (dextran hydrolysate ladder) pioneered by Waters and the National Institute for Bioprocessing Research and Training (Dublin, Ireland).[13] Some isobaric glycans were not well resolved with the chromatographic method used; therefore, peak heights were used for quantification.

Digestion of the native NIST mAb using resin-bound SMART Digest trypsin produced glycopeptides with the same glycans observed from the PNGase F-treated, fluorescently labeled sample. However, the retention times of the glycopeptides were shifted to longer elution times, as a result of the additional retentivity produced by the tryptic peptide of interest EEQYNSTYR (**Fig. 2**). The glycan elution order (as attached glycopeptides) was the same as the fluorescently labeled glycans (**Fig. 3**).

Table 2 details the PNGase F-released, fluorescently labeled glycans and the glycopeptides that were found. Negative mode electrospray ionization was used for the glycopeptide analysis, as a result of adduct formation with ammonia and in-source decay, which was problematic in positive-mode electrospray ionization. Positive-mode

T A B L E 2

List of detected released glycans and glycopeptides

| Glycan name | GU value | Formula | RapiFluor-labeled released glycans | | | | Glycopeptides with glycans attached to EEQYNSTYR | | | |
			z	m/z	R.T., min	Intensity, cts	z	m/z	R.T., min	Intensity, cts
F(6)M2	3.740	C34H58N2O25	1	1206.5017	11.84	1.30E + 03				
M3	4.380	C34H58N2O26	1	1222.5077	11.94	9.87E + 02	−2	1039.4805	21.84	6.38E + 02
F(6)A1	5.313	C48H81N3O35	2	786.3249	11.91	1.97E + 05	−2	1213.9806	24.73	1.10E + 04
F(6)A2 = G0F	5.818	C56H94N4O40	2	887.8733	13.13	6.04E + 05	−2	1315.5133	25.24	3.00E + 05
F(6)A1G(4)1	6.218	C54H91N3O40	2	867.3481	14.41	3.75E + 04	−2	1294.9906	26.36	1.23E + 04
F(6)A2[6]G(4)1 = G1F	6.563	C62H104N4O45	2	968.8922	15.29	6.64E + 05	−2	1396.5380	26.43	1.87E + 05
F(6)A2[6]BG(4)1 = G1F + G	6.853	C70H117N5O50	2	1070.4351	16.23	3.03E + 03	−2	1498.0931	26.77	1.12E + 03
F(6)A3G(4)1	6.907	C70H117N5O50	2	1070.4351	16.47	3.10E + 03	−2	1498.0762	27.07	6.71E + 02
F(6)M5A1	6.910	C60H101N3O45	2	948.3731	16.40	1.63E + 03	−2	1376.0386	27.62	1.62E + 03
F(6)A1G(4)1Ga1	7.120	C60H101N3O45	2	948.3865	16.81	1.40E + 04	−2	1376.0223	27.79	3.81E + 03
F(6)A2G(4)2 = G2F	7.375	C68H114N4O50	2	1049.9226	17.52	1.57E + 05	−2	1477.5474	27.79	2.30E + 04
F(6)A2BG(4)2 = G2F + G	7.636	C76H127N5O55	2	1151.4600	18.23	2.88E + 03	−2	1579.1239	28.02	1.28E + 03
F(6)M4A1G(4)1Ga(3)1	7.865	C66H111N3O50	2	1029.3986	18.60	2.24E + 03	−2	1457.0430	28.26	1.10E + 03
F(6)A1G(4)1Sg(6)1	7.971	C65H108N4O49	2	1020.8996	19.07	2.85E + 04	−2	1448.5326	29.15	1.98E + 03
F(6)A2G(4)2Ga(3)1	8.246	C74H124N4O55	2	1130.9449	19.48	1.87E + 04	−2	1558.5798	28.91	2.54E + 03
A2G(4)2S(6)1	8.321	C73H121N5O54	2	1122.4544	19.51	4.07E + 03	−2	1550.0696	29.45	2.46E + 02
F(6)A2G(4)1Sg1 iso	8.353	C73H121N5O54	2	1122.4308	19.88	5.88E + 03	−2			
F(6)M4A1G(4)1Sg(6)1	8.660	C71H118N4O54	2	1101.9164	20.76	1.52E + 03	−2	1529.5623	30.13	3.27E + 02
F(6)A2G(4)2Ga(3)2	9.108	C80H134N4O60	2	1211.9672	21.30	1.03E + 04	−2	1639.5984	30.10	1.20E + 03
M5A1G(4)1Sg(6)1	9.276	C71H118N4O55	2	1109.9452	21.94	8.67E + 02	−2	1537.4849	30.71	1.76E + 02
F(6)M5A1G(4)1Sg(6)1	9.569	C77H128N4O59	2	1182.9609	22.45	7.55E + 02	−3	1073.3810	31.25	2.42E + 02
F(6)A2G(4)2Ga(3)1Sgz(6)1	9.877	C85H141N5O64	2	1284.5016	23.02	6.94E + 03				

cts, Counts; GU, glucose unit; m/z, mass-to-charge ratio; R.T., retention time (in minutes); z, charge.

electrospray ionization was used for the PNGase F-released, fluorescently labeled glycans, as a result of the enhanced ionization efficiency of the Waters RapiFluor label.

The analysis of digest completion was investigated for the glycopeptide method (**Fig. 4**). Interestingly, miscleavage products for the tryptic digest were not detected. The data in **Fig. 5** correlate the intensities of the molecular ions for the +/−2 charge states between the labeled glycans and glycopeptides.

For the resin-bound SMART Digest, all of the mAb appear to be digested to yield the desired tryptic (glyco)peptide. The use of this type of digest presents several advantages, including reduced digestion time and dramatic lowering of miscleavages. This latter effect is likely, at least in part, a result of the resin-bound nature of the protease, lowering the possibility of trypsin–trypsin interactions (Perfinity Biosciences; Application Note, http://www.perfinity.com/flash-digest-applications-by-type).[14]

DISCUSSION

In this work, we have demonstrated a highly efficient digestion method to produce predictable glycopeptides with no detected miscleavages. The developed resin-bound digestion method correlates well with the traditional method for glycan analysis *via* PNGase F-released glycans. Efficient digestion of the protein at the site of glycosylation with minimal miscleavages allows the use of resulting glycopeptides to quantify glycosylations in site-specific fashion. With the appropriate standards, this method could allow for site-specific absolute quantification of glycosylations on proteins.

Whereas increase of the method sensitivity is ongoing, the current signal intensity is 1000-fold above the noise level and would limit miscleavage products, if they exist, to <0.1% in abundance relative to the desired tryptic peptide. This should allow for absolute quantification of the glycosylation on the mAb with the appropriate standards.
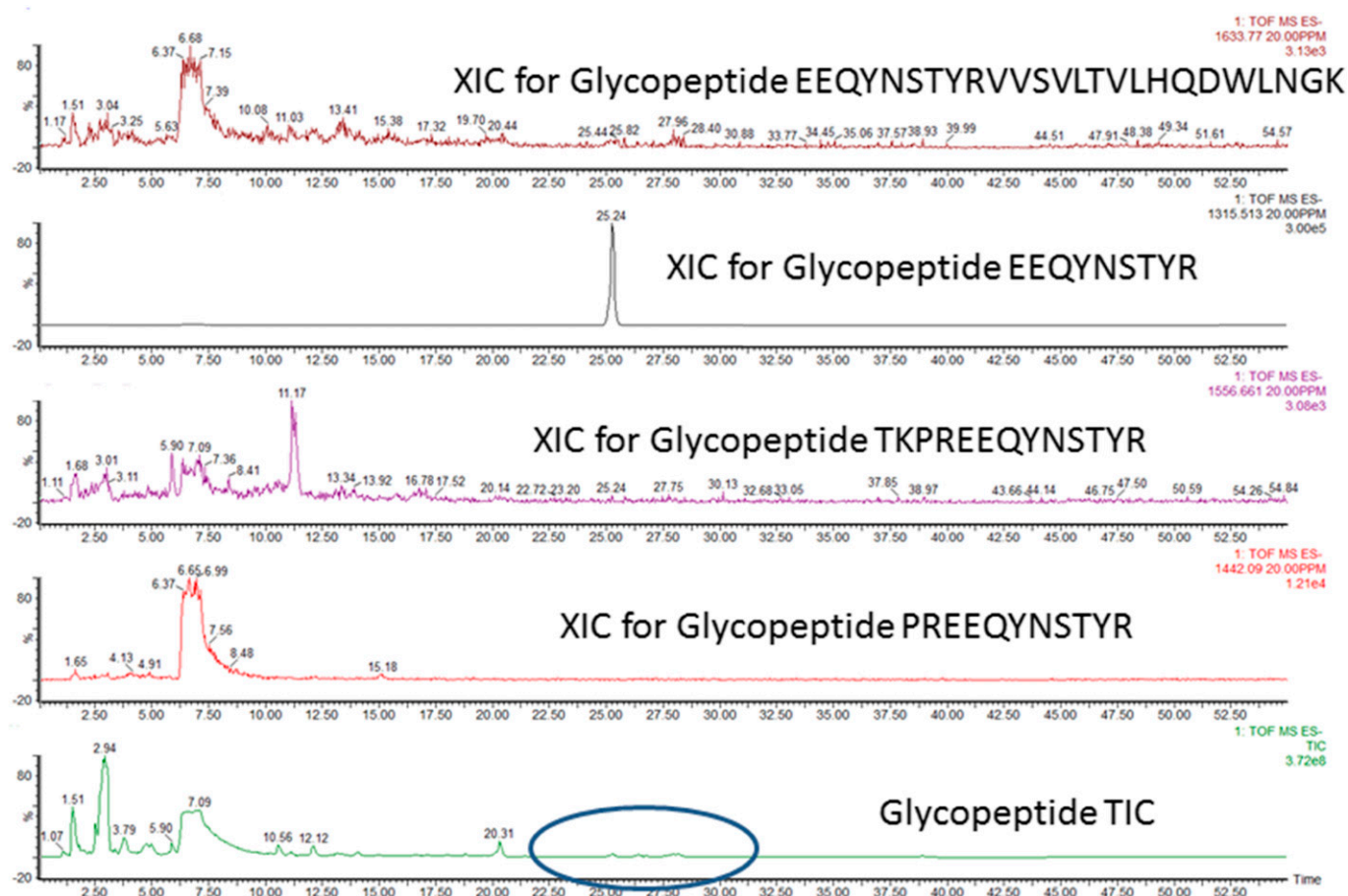

FIGURE 4

Extracted ion chromatograms (XIC) for potential NIST mAb glycopeptide miscleavages. The circled region of the TIC is the elution range for the glycopeptides. Extracted ion chromatograms for the potential miscleavage products were attempted for the −3 and −2 charge states. The intensity of the glycopeptide EEQYNSTYR, with the glycan G0F attached, is roughly 300,000 counts. Noise levels in the other traces in the same retention time range are between 100 and 300 counts.
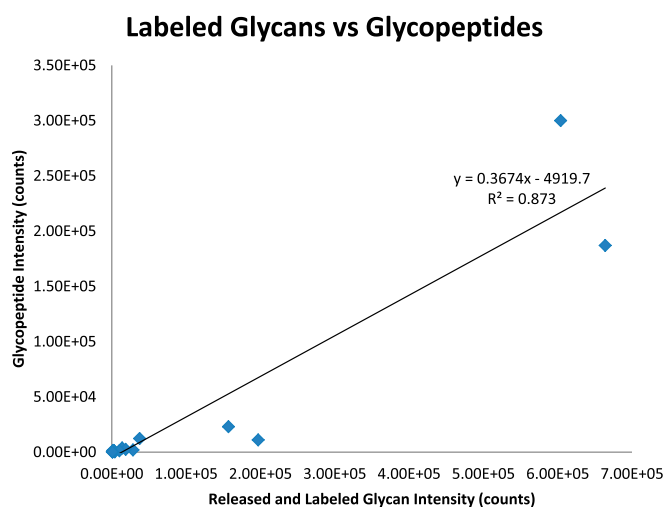
Labeled Glycans vs Glycopeptides

$$y = 0.3674x - 4919.7$$
$$R^2 = 0.873$$

FIGURE 5

Intensity correlation between labeled glycans and glycopeptides. Intensity levels from free glycan *vs.* glycopeptide are plotted. Good correlation exists between the quantification of glycopeptides and labeled glycans. Some of the anomalies may exist as a result of chromatography issues. The chromatography method did not fully resolve some isobaric glycans, thereby complicating quantification. Multiple charge states for the same analyte are not summed.

Any other glycoprotein, if it could be digested with similar efficiency, could have its glycosylation quantified in a similar site-specific fashion. Moreover, as a result of the conserved nature of the Fc glycosylation site in human antibodies, a peptide with the sequence EEQYNSTYR with a glycan library attached to the asparagine could be a widely applicable standard for quantification of human mAb glycosylation.

This method relies on the ability to produce a single peptide from each glycosylation site of a glycoprotein. The proteins that do not have sequences amenable to tryptic digestion at the site(s) of glycosylation would have to be analyzed using other proteases. The development of other thermally stable, resin-bound proteases is underway, and the method we have described should be applicable.

We believe that this analytical method, coupling resin-bound digestion with LC/MS for glycoprotein analysis, has considerable advantages over current procedures and can help provide insight in glycobiology.

be identified in this article to specify adequately the experimental procedure. Such identification does not imply approval, endorsement, or certification by USP of a particular brand or product nor does it imply that the equipment, instrument, vendor, or material is necessarily the best available for the purpose or that any other brand or product was judged to be unsatisfactory or inadequate. All product names, logos, and brands are property of their respective owners.

REFERENCES

1. Varki A. Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* 1993;3:97–130.
2. Hossler P, Khattak SF, Li ZJ. Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* 2009;19: 936–949.
3. Mariño K, Bones J, Kattla JJ, Rudd PM. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol* 2010;6:713–723.
4. An Y, Cipollo JF. An unbiased approach for analysis of protein glycosylation and application to influenza vaccine hemagglutinin. *Anal Biochem* 2011;415:67–80.
5. Thaysen-Andersen M, Packer NH, Schulz BL. Maturing glycoproteomics technologies provide unique structural insights into the N-glycoproteome and its regulation in health and disease. *Mol Cell Proteomics* 2016;15:1773–1790.
6. Morelle W, Michalski JC. Analysis of protein glycosylation by mass spectrometry. *Nat Protoc* 2007;2:1585–1602.
7. Lauber MA, Koza SM, Chambers EE. Comprehensive characterization of the N and O-linked glycosylation of a recombinant human EPO. *Application Notes*, 720005462EN. Milford, MA: Waters. Retrieved August 10, 2017, from http://www.waters.com/waters/library.htm?locale=en_US&lid=134855745.
8. An HJ, Froehlich JW, Lebrilla CB. Determination of glycosylation sites and site-specific heterogeneity in glycoproteins. *Curr Opin Chem Biol* 2009;13:421–426.
9. Šlechtová T, Gilar M, Kalíková K, Tesařová E. Insight into trypsin miscleavage: comparison kinetic constants of problematic peptide sequences. *Anal Chem* 2015;87:7636–7643.
10. Samonig M, Schwahn A, Cook K, Oliver M, Swart R. SMART Digest compared to classic in-solution digestion of rituximab for in-depth peptide mapping characterization. *Application Note No. 1159*. Waltham, MA: Thermo Fisher Scientific. Retrieved August 10, 2017, from https://tools.thermofisher.com/content/sfs/brochures/AN-1159-SP-SMART-Digest-Peptides-AN72141-EN.pdf.
11. Meyer K, O'Grady J, Poe D. Enhanced trypsin digestion for improved biomarker sensitivity and peptide identification. *Application Notes*. West Lafayette, IN: Perfinity Biosciences. Retrieved August 10, 2017, from http://www.perfinity.com/flash-digest-applications-by-type.
12. National Institute of Standards & Technology. Report of investigation. Reference material 8671. NISTmAb, humanized IgG1κ monoclonal antibody. Lot No. 14HB-D-0002. Gaithersburg, MD: U.S. Department of Commerce. Retrieved August 10, 2017, from https://www-s.nist.gov/srmors/certificates/8671.pdf.
13. A systematic approach to glycan analysis using HILIC-UPLC and an online database of standardized values. *Technology Brief*. Milford, MA: Waters. Retrieved August 10, 2017, from http://www.waters.com/webassets/cms/library/docs/720004203en.pdf.
14. Perfinity: controlling the major factors impacting trypsin digestion. *Application Notes*. West Lafayette, IN: Perfinity Biosciences. Retrieved August 10, 2017, from http://www.perfinity.com/flash-digest-applications-by-type.